7th Symposium of the European Association for Research in Transportation conference – hEART 2018
5-7 September 2018, Athens, Greece

Comparing theoretical pedestrian simulation models with data driven techniques

George Kouskoulis [a], Constantinos Antoniou [b], Ioanna Spyropoulou [c]

[a] *Ph.D. Candidate, School of Rural and Surveying Engineering, National Technical University of Athens, Iroon Polytechneiou Str. 9, Zographos, Athens, 157 80, Greece*
[b] *Professor, Chair of Transportation Systems Engineering, Technical University of Munich, Arcisstraße 21, Munich, D-80333, Germany*
[c] *Assistant Professor, School of Rural and Surveying Engineering, National Technical University of Athens, Iroon Polytechneiou Str. 9, Zographos, Athens, 157 80, Greece*

Keywords: Pedestrian modeling, data driven, loess, social force model, data noise reduction

1. Introduction

Following up on the continuous widespread availability of data and computational advances, data driven modeling has been increasingly gaining researchers' interest over the last decades (e.g. Antoniou et al., 2013; Papathanasopoulou and Antoniou, 2015). Methods and techniques have been developed on clustering, classifying and regressing data with no necessary explicit a priori knowledge of model parameters' relationships. These methods are not subject to parametric limitations and are thus wider applicable.

On the other hand, theoretical models provide a straight mathematical framework, while relating model parameters based on logical principles. They may be categorized as follows (Kouskoulis and Antoniou, 2017):

- Cellular automata
- Social force
- Lattice gas theory

Cellular automata models rely on space discretization, social force on the interactive forces that are acted among moving pedestrians and lattice gas on the drift strength (strength of pedestrian flow). Both cellular automata and lattice gas models compute the possibility a pedestrian's next step based on the surroundings (obstacles and agents). Thus, the basic principle on modelling pedestrian movement lies on the positions of the adjacent pedestrians/objects and their impact on the examined agents. This is directly represented in the dynamics of the social force models. Even more, social force models outweigh on computational requirements.

The aim of this paper is to provide a comparison between data driven and theoretical pedestrian simulation models. As a study case, the locally weighted regression and the social force model are utilized, respectively, using data-sets collected by the first author.

2. Methodological framework

The key methodological steps of the research approach include:

- Data collection – video recording
- Data filtering – trajectory smoothing
- Model design implementation
- Model comparison – cross-validation method

The primary step of the methodology process is to collect the appropriate data for the experiment. Data collection techniques include several tools, such as radars, lasers, cameras, sensors, GPS, and exhibit different strengths and weaknesses, and different accuracy levels (Antoniou et al., 2011). In the current experimental design, video cameras are employed for recording pedestrian trajectories. As initial trajectory data include noise, data smoothing is applied. Subsequently to noise elimination, a theoretical pedestrian simulation model and a data driven technique are implemented. Last, the two models are compared through an appropriate goodness-of-fit indicator.

3. Case study setup

3.1. Data collection and processing

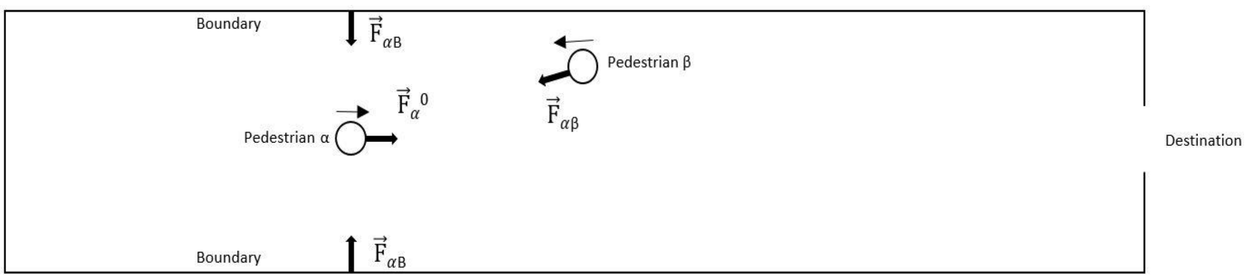
Trajectory data have been collected from the (aboveground) platforms of Moschato metro station in Athens [Figure 1 (a)] in the morning (8:30 am) and a shopping mall (indoor) in Athens [Figure 1 (b)] in the evening (7:40 pm) (both in non-working days). With the aid of the tracking software named “Tracker – Video Analysis and Modeling Tool” (version 4.11.0) video recordings were extracted to pedestrian trajectories data, with the pixel coordinates were then transformed to real world coordinates with the use of appropriate photogrammetric tools and software (Kalisperakis et al., 2006; and Douskos et al., 2009). A combination of Kalman Filter (UKF) and moving average extensions were employed in order to eliminate data noise.

(a) **(b)**

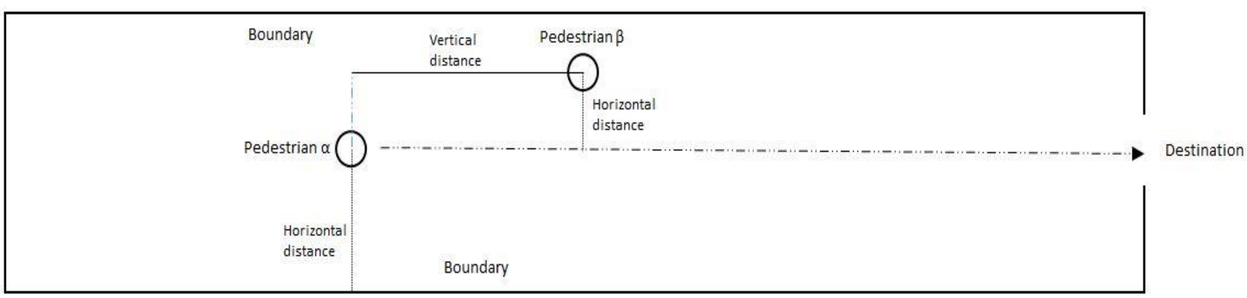
FIGURE 1 Collecting data places. (a) Metro station, (b) Mall

3.2. Model setup

Two modeling approaches are employed: (i) a theory-derived social force model, and (ii) a data driven loess technique. A visual overview of their modeling assumptions is provided in Figure 2.



(a) Social force description



(b) Data driven model setup

FIGURE 2 Pedestrian movement representation.

Helbing and Molnár (1995) presented the social force model as an effort for simulating pedestrian kinematics. The model has since been employed from widely applied simulation software [e.g. Vissim (PTV, 2016) and SimWalk (Zainuddin et al., 2009)]. The model's theory is that as an agent walks he/she receives forces from his/her surroundings that coerce him/her to amend his/her velocity, similarly to the forces in fluid molecules. Social forces are distinguished in attractive and repulsive. The force that affects pedestrian movement ($\vec{F_\alpha}(t)$) is the sum of the aforementioned forces (equation (1)).

$$\vec{F}_\alpha = \vec{F}_\alpha^0 + (\vec{F}_{\alpha\beta} + \vec{F}_{\alpha \mathrm{B}}) + \vec{F}_{\alpha i} \quad (1)$$

where $\vec{F}_\alpha^0$ are the attractive forces acted on pedestrian a from the destination, while $\vec{F}_{\alpha\beta}$ the repulsive forces from other pedestrians β and $\vec{F}_{\alpha \mathrm{B}}$ from boundaries B and $\vec{F}_{\alpha i}$ the attractive forces from other agents or objects i.

Helbing and Johansson (2009) differentiated the social force model pedestrian's repulsive forces equation (9) by importing interaction strength (A) and interaction range (B) parameters.

As an outcome of social force model pedestrians alter their velocity, as they walk, based on six predictors (equation (2)).

$$\vec{u}_\alpha(t) = f(\vec{u}_\alpha(t-1), \vec{r}_{\alpha\beta}, \vec{r}_{\alpha \mathrm{B}}, \vec{r}_\alpha^{\,k}, \vec{r}_{\alpha i}, \vec{u}_\beta) \qquad (2)$$

where $\vec{u}_\alpha(t-1)$ is pedestrian velocity in the previous time step, $\vec{r}_{\alpha\beta}$, $\vec{r}_{\alpha \mathrm{B}}$, $\vec{r}_{\alpha i}$ the distances between pedestrians/obstacles and $\vec{r}_\alpha^{\,k}$ the distance to the destination.

Locally weighted regression (LOESS, Cleveland 1979) is a widely applied data driven method employed for predicting and regression analysis that fits data points, based on smoothing technique and weighted least squares. In practice, the method performs efficiently in one predictor modeling, while its performance is reduced (curse of dimensionality) as the number of predictors is increased (Cleveland and Devlin, 1988; Cleveland et al., 1988).

To achieve comparable results the same predictors/parameters are implemented in both models. However, as -due to implementation details- loess predictors are limited to three, the following variables are considered i) pedestrian velocity in the previous time step ($\vec{u}_\alpha(t-1)$, ii) distance between the examined agent and the pedestrians triggering repulsive effects ($\vec{r}_{\alpha\beta}$) and iii) distance between the examined agent and space boundaries ($\vec{r}_{\alpha \mathrm{B}}$).

4. Model comparison

A full cross-validation pattern is employed in order to compare the two models. A dataset is separated into k parts (folds). In each run, one of the k folds is used as the testing set while all the others as the training set. The model is applied in the training set by altering its parameters and selecting the parameter values combination that minimizes the total error (goodness of fit measure). Subsequently the model with the selected parameter values is applied to the testing dataset capturing its validity (the error in the testing set is computed). After every of the k folds has become a testing, set the cross-validation process is complete and the total error of the testing sets is computed. The model type with the lowest error value implies to be more appropriate for simulating the phenomenon.

In the specific experiment five datasets are used and five different training/testing cycles are performed (5-fold cross-validation). As the datasets are not of the same size, they are merged and then divided in to five equal sized parts. Each time four of the datasets are used for training, and the remaining one for testing. In the case of loess, the training part involves the selection of the optimal span (between 0.1 and 0.9, to avoid extreme values) and degree (1 or 2), while for the social force model, it revolves around the determination of the desired speed distribution (mean and standard deviation) and the maximum acceptable speed (coefficient), as they have been indicated as the critical model parameters from sensitivity analysis that has been conducted in a previous step. Table 1 presents the results. Results indicate that in this particular application, the loess model seems to provide superior performance. Despite the limitations, related to the opacity and the (lack of) interpretability of the model, it seems a promising avenue for model development, when model fit is the primary concern. The model seems to provide rather accurate estimations.

TABLE 1 Model performance comparison

Training folds/Testing folds	Model type			
	Social force model		Loess	
RMSPE	Training folds	Testing folds	Training folds	Testing folds
1,2,3,4/5	17.45%	16.44%	7.92%	2.95%
1,2,3,5/4	18.08%	18.94%	7.95%	3.16%
1,2,4,5/3	19.70%	9.15%	7.96%	2.95%
1,3,4,5/2	12.52%	31.71%	6.20%	6.54%
2,3,4,5/1	16.25%	28.47%	6.32%	6.91%

5. Discussion

This paper provides a contribution towards exploring data-driven techniques' efficacy on pedestrian modeling. Following model training and testing, simulation results indicated, through the RMSPE index, significantly higher (better) performance for loess. The results of this study demonstrate that loess is a very promising approach for pedestrian simulation, as it is expected to provide accurate results. Loess higher simulation performance, lower computational time requirement and "mathematical" simplicity make it a suitable method for pedestrian simulation. A limitation is the small number of features that can be currently considered. Therefore, future research includes the employment of additional data driven techniques (e.g. neural networks, support vector machines) and goodness-of-fit measures that will allow an in-depth comparative analysis.

References

Antoniou, C., R. Balakrishna and H. N. Koutsopoulos (2011). A synthesis of emerging data collection technologies and their impact on traffic management applications. European Transport Research Review 3, pp. 139-148.

Antoniou, C., H. N. Koutsopoulos and G. Yannis (2013). Dynamic Data-Driven Local Traffic State Estimation and Prediction. Transportation Research Part C: Emerging Technologies, Vol. 34, pp. 89-107.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association, Vol. 74, No. 368, pp. 829–836.

Cleveland, W. S. and S. J. Devlin (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association, Vol. 83, No. 403, pp. 596-610.

Cleveland, W. S., S. J. Devlin and E. Grosse (1988). Regression by Local Fitting: Methods, Properties, and Computational Algorithms. Journal of Econometrics, Vol. 37, pp. 87-114.

Douskos, V., L. Grammatikopoulos, I. Kalisperakis, G. Karras, E. Petsa (2009). FAUCCAL: an open source toolbox for fully automatic camera calibration, XXII CIPA Symposium on Digital Documentation, Interpretation & Presentation of Cultural Heritage, Kyoto, Japan.

Helbing, D. and P. Molnár (1995). Social force model for pedestrian dynamics, Physical Review E, Vol. 51, pp. 4282-4286.

Helbing, D. and A. Johansson (2009). Pedestrian, Crowd and Evacuation Dynamics, Encyclopedia of Complexity and Systems Science, pp. 6476-6495, Print ISBN 978-0-387-75888-6, Online ISBN 978-0-387-30440-3, Publisher Springer New York

Kalisperakis, I., L. Grammatikopoulos, E. Petsa and G. Karras (2006). An open-source educational software for basic photogrammetric tasks, FIG-ISPRS-ICA International Symposium on Modern Technologies, Education & Professional Practice in Geodesy & Related Fields, Sofia, Bulgaria, pp. 581-586.

Kouskoulis, G. and C. Antoniou (2017). A systematic review of pedestrian simulation models with a focus on emergency situations. Transportation Research Record: Journal of the Transportation Research Board (in press).

Papathanasopoulou, V. and C. Antoniou (2015). Towards data-driven car-following models. Transportation Research Part C, Vol. 55, pp. 496-509.

PTV VISSIM 7 User Manual. PTV AG, 2015.

Zainuddin, Z., K. Thinakaran and I. Abu-Sulyman (2009). Simulating the Circumambulation of the Ka'aba using SimWalk. European Journal of Scientific Research, Vol. 38, No. 3, pp. 454-464.